BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 3: e37223
doi: 10.3897/biss.3.37223

OPEN ACCESS

Conference Abstract

The Semantic Field Book Annotator

Lise Stork[‡], Andreas Weber[§], Katherine Wolstencroft[‡]

‡ Leiden Institute of Advanced Computer Science, Leiden, Netherlands
§ University of Twente, Enschede, Netherlands

Corresponding author: Lise Stork (l.stork@liacs.leidenuniv.nl)

Received: 12 Jun 2019 | Published: 19 Jun 2019

Citation: Stork L, Weber A, Wolstencroft K (2019) The Semantic Field Book Annotator. Biodiversity Information Science and Standards 3: e37223. https://doi.org/10.3897/biss.3.37223

Abstract

Biodiversity research expeditions to the globe's most biodiverse areas have been conducted for several hundred years. Natural history museums contain a wealth of historical materials from such expeditions, but they are stored in a fragmented way. As a consequence links between the various resources, e.g., specimens, illustrations and field notes, are often lost and are not easily re-established.

Natural history museums have started to use persistent identifiers for physical collection objects, such as specimens, as well as associated information resources, such as web pages and multimedia. As a result, these resources can more easily be linked, using Linked Open Data (LOD), to information sources on the web. Specimens can be linked to taxonomic backbones of data providers, e.g., the Encyclopedia Of Life (EOL), the Global Biodiversity Information Facility (GBIF), or publications with Digital Object Identifiers (DOI).

For the *content* of biodiversity expedition archives, (e.g. field notes), no such formalisations exist. However, linking the specimens to specific handwritten notes taken in the field can increase their scientific value. Specimens are generally accompanied by a label containing the location of the site where the specimen was collected, the collector's name and the classification. Field notes often augment the basic metadata found with specimens with important details concerning, for instance, an organism's habitat and morphology. Therefore, inter-collection interoperability of multimodal resources is just as important as intra-collection interoperability of unimodal resources.

© Stork L et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

The linking of field notes and illustrations to specimens entails a number of challenges: historical handwritten content is generally difficult to read and interpret, especially due to changing taxonomic systems, nomenclature and collection practices. It is vital that:

1. the content is structured in a similar way as the specimens, so that links can more easily be re-established either manually or in an automated way;
2. for consolidation, the content is enriched with outgoing links to semantic resources, such as Geonames or Virtual International Authority File (VIAF); and
3. this process is a transparent one: how links are established, why and by whom, should be stored to encourage scholarly discussions and to promote the attribution of efforts.

In order to address some of these issues, we have built a tool, the Semantic Field Book Annotator (SFB-A), that allows for the direct annotation of digitised (scanned) pages of field books and illustrations with Linked Open Data (LOD). The tool guides the user through the annotation process, so that semantic links are automatically generated in a formalised way. These annotations and links are subsequently stored in an RDF triplestore.

As the use of the Darwin Core standard is considered best practice among collection managers for the digitisation of their specimens, our tool is equipped with an ontology based on Darwin Core terms, the NHC-Ontology, which extends the Darwin Semantic Web (DSW) ontology. The tool can annotate any image, be it an image of a specimen with a textual label, an illustration with a textual label or a handwritten species description. Interoperability of annotations between the various resources within a collection is therefore ensured. Terms in the ontology are structured using OWL web ontology language. This allows for more complex tasks such as OWL reasoning and semantic queries, and facilitates the creation of a richer knowledge base that is more amenable to research.

Keywords

semantic annotation, field books, ontologies, linked data

Presenting author

Lise Stork

Presented at

Biodiversity_Next 2019